Journal of
Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Risk analysis, ideal observers, and receiver operating characteristic curves for tasks that combine detection and estimation

Eric Clarkson

SPIE.

Eric Clarkson, "Risk analysis, ideal observers, and receiver operating characteristic curves for tasks that combine detection and estimation," *J. Med. Imag.* **6**(1), 015502 (2019), doi: 10.1117/1.JMI.6.1.015502.

Journal of Medical Imaging 6(1), 015502 (Jan–Mar 2019)

Risk analysis, ideal observers, and receiver operating characteristic curves for tasks that combine detection and estimation

Eric Clarkson*
University of Arizona, College of Optical Sciences, Tucson, Arizona, United States

Abstract. Previously published work on joint estimation/detection tasks has focused on the area under the estimation receiver operating characteristic (EROC) curve as a figure of merit (FOM) for these tasks in imaging. Another FOM for these joint tasks is the Bayesian risk, where a cost is assigned to all detection outcomes and to the estimation errors, and then averaged over all sources of randomness in the object ensemble and the imaging system. Important elements of the cost function, which are not included in standard EROC analysis, are that the cost for a false positive depends on the estimate produced for the parameter vector, and the cost for a false negative depends on the true value of the parameter vector. The ideal observer in this setting, which minimizes the risk, is derived for two applications. In the first application, a parameter vector is estimated only in the case of a signal present classification. For the second application, parameter vectors are estimated for either classification, and these vectors may have different dimensions. In both applications, a risk-based estimation receiver operating characteristic curve is defined and an expression for the area under this curve is given. It is also shown that, for some observers, this area may be estimated from a two alternative forced choice test. Finally, if the classifier is optimized for a given estimator, then it is shown that the slope of the risk-based estimation receiver operating characteristic curve at each point is the negative of the ratio of the prior probabilities for the two classes. [DOI: 10.1117/1.JMI.6.1.015502]

Keywords: image quality assessment; probabilistic and statistical methods; quantification and estimation; receiver operating characteristic analysis; Bayesian ideal observers; estimation receiver operating characteristic curves.

Paper 18108RR received May 27, 2018; accepted for publication Feb. 5, 2019; published online Feb. 22, 2019.

1 Introduction

Some imaging tasks involve the detection of a signal combined with the estimation of parameters. One possibility is that we want to estimate parameters associated with the signal, such as its location, shape, size, or composition. This situation is described in application I below. Another possibility is that we want to estimate parameters associated with the imaged object whether a signal is present or not, such as voxel values for a three-dimensional reconstruction of the object. This situation is discussed in application II below. We may also want to perform both estimation tasks. This situation is also covered in application II.

In order to optimize an imaging system for these joint tasks, we need a figure of merit (FOM) that takes into account the value of the data for each component, the detection component and the estimation component. One such FOM is the area under the estimation receiver operating characteristic (EROC) curve, which was introduced by the author in previous papers[1,2] as a generalization of the localization ROC curve. The EROC curve and generalizations have also been studied by other researchers as a practical method for quantifying the performance of imaging systems on joint tasks.[3,4] Another FOM for this kind of task is the Bayesian risk, where a cost is assigned to all possible outcomes and then averaged over all possible objects and noise realizations.

For each application, we derive the Bayesian ideal observer, which is a mathematical algorithm that performs the task and minimizes the risk in the process. We also define a risk-based estimation receiver operating characteristic (RB-EROC) curve for each application that is a generalization of standard ROC and EROC curves. We then derive expressions for the areas under these curves and show how, for certain suboptimal observers, these areas can be estimated from a two alternative forced choice (2AFC) test. We also show that, for a class of suboptimal observers, the slope of the RB-EROC curve at each point is determined by the prior class probabilities corresponding to that point.

An important aspect of the cost functions used here is that we allow the dependence of the cost of each decision to depend on the parameter vector or an estimate of it. For example, in application I, the cost of a false positive can depend on the estimated parameter vector since this vector will affect future actions based on the misclassification. Similarly, for a false negative, the cost is allowed to depend on the true parameter vector since the value of this vector may affect the consequences of this misclassification for the patient. The cost for a true positive classification depends on both the true parameter and its estimate as in the standard EROC analysis (which uses utility instead of cost). Finally, the cost for a true negative is a constant since there is no true parameter vector or estimate in that case, which is also the case for standard EROC analysis. As we will see below, similar dependences of the costs on true parameter vectors and their estimates for the various decision outcomes are also allowed in application II.

*Address all correspondence to Eric Clarkson, E-mail: clarksoneric@mac.com

2329-4302/2019/$25.00 © 2019 SPIE

2 Application I: Conditional Estimation

In this task, the observer uses the data vector $\mathbf{g}$ to decide whether this vector was the result of imaging a member of the signal absent class, which is hypothesis H_0, or a member of the signal present class, which is hypothesis H_1. All data vectors are assumed to be in a data space $\mathcal{D}$, which is a subset of $\mathbb{R}^M$. If the observer decides on signal present then an estimate $\hat{\boldsymbol{\theta}}$ for the parameter vector $\boldsymbol{\theta}$ must be computed. Thus, the estimation task is conditional on the outcome of the detection task. For the detection component of the joint task, we have a cost assigned to each outcome: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The probability distribution function (PDF) of the data vector $\mathbf{g}$ when the signal is present is $pr(\mathbf{g}|1)$. When the signal is absent, the PDF for the data vector is $pr(\mathbf{g}|0)$. We then have the elements of the cost matrix

$$\begin{bmatrix} C_{00} & C_{01}[\boldsymbol{\theta}] \\ C_{10}[\hat{\theta}] & C_{11}[\hat{\theta},\boldsymbol{\theta}] \end{bmatrix},$$

where C_{00} is the cost associated with a TN outcome, $C_{10}(\hat{\boldsymbol{\theta}})$ is the cost for an FP, $C_{11}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})$ the cost for a TP, and $C_{01}(\boldsymbol{\theta})$ is the cost for an FN. We assume that $C_{10}(\hat{\boldsymbol{\theta}}) > C_{00}$. In other words, it costs less to get a TN classification than to get an FP. Since a TN outcome will usually lead to no further intervention, and no harm resulting from this action, whereas an FP outcome will usually lead to further intervention, this assumption seems reasonable. Later we will see at what point in the calculation this assumption is used and comment on what the decision procedure would be without it. For TP outcome, we have a nonnegative cost function $C_{11}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})$, which assigns a cost to the estimate $\hat{\boldsymbol{\theta}}$ when the true parameter vector is $\boldsymbol{\theta}$. For FN outcomes, the cost to the patient for a misdiagnosis depends on the true parameter vector. The probability that the signal is present is P_1 and the probability that it is absent is P_0. These two numbers are assumed to be positive and to sum to unity. This cost matrix is discussed in Refs. 5 and 6, and an equivalent utility formulation is discussed in Ref. 3. Later in this section, we will indicate where the calculations given here deviate from those references.

In this application, we assume that the parameters to be estimated are associated with the signal present cases only. In medical imaging, for example, the parameters contained in $\boldsymbol{\theta}$ may describe the size, shape, location, and other factors associated with a tumor. For TP cases, the cost depends on how close the estimate $\hat{\boldsymbol{\theta}}$ is to the true vector $\boldsymbol{\theta}$. For FP cases, the value of the estimate $\hat{\boldsymbol{\theta}}$ will affect further studies and treatments, and will therefore affect the cost to the patient. For FN cases, the true vector $\boldsymbol{\theta}$ of an undetected tumor will surely affect future costs to the patient as the cancer progresses. Finally, for TN cases, there is no cancer and no further studies or treatments are performed so the cost in this case is a constant and may in fact be zero. In medical imaging writing, explicit expressions for these costs are a difficult task, but these costs do exist. Therefore, there is an optimal observer that minimizes the average cost, i.e., the risk. The closer we can come to implementing this observer, even with approximate or idealized cost expressions, the lower the risk will be. This ideal observer may also offer insight into strategies for improving observer performance. For example, as we will see, the ideal observer performs the estimation task first and then uses the estimate in the classification task.

We write the signal-present PDF in terms of a conditional PDF $pr(\mathbf{g}|\boldsymbol{\theta})$ determined by the imaging system and nuisance parameters in the object, and a prior PDF $pr(\boldsymbol{\theta})$ on the parameters of interest:

$$pr(\mathbf{g}|1) = \int_\Theta pr(\mathbf{g}|\boldsymbol{\theta})pr(\boldsymbol{\theta})\mathrm{d}^L\theta. \tag{1}$$

In this integral, Θ is the domain of the prior PDF $pr(\boldsymbol{\theta})$ and is assumed to be a subset of $\mathbb{R}^L$. Let $\mathcal{A}$ be the region in the data space $\mathcal{D}$ that consists of the data vectors that the observer declares to be in the signal absent class. Similarly, let $\mathcal{P}$ be the region in the data space $\mathcal{D}$ that consists of the data vectors that the observer declares to be in the signal present class. We make the usual assumptions that $\mathcal{A}\cap\mathcal{P}=\varnothing$ and $\mathcal{A}\cup\mathcal{P}=\mathcal{D}$. Now we may write the risk in integral form as

$$\begin{aligned} C = {} & P_0\left\{\int_\mathcal{A} C_{00}pr(\mathbf{g}|0)\mathrm{d}^M g + \int_\mathcal{P} C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g})]pr(\mathbf{g}|0)\mathrm{d}^M g\right\} \\ & + P_1\left\{\left\langle\int_\mathcal{A} C_{01}[\boldsymbol{\theta}]pr(\mathbf{g}|\boldsymbol{\theta})\mathrm{d}^M g\right\rangle_{\boldsymbol{\theta}}\right. \\ & \left. + \left\langle\int_\mathcal{P} C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]pr(\mathbf{g}|\boldsymbol{\theta})\mathrm{d}^M g\right\rangle_{\boldsymbol{\theta}}\right\}. \end{aligned} \tag{2}$$

The first two terms correspond to the FP and TN cases, while the last two terms are from the FN and TP cases. At this point in Ref. 5, the costs C_{10} and C_{01} are assumed to be constant and a special form for $C_{11}[\hat{\boldsymbol{\theta}},\boldsymbol{\theta}]$ is used. We will not be making those restrictions in what follows.

2.1 *Ideal Observer for Application I*

To compute the observer that minimizes the risk, the Bayesian ideal observer, we change the order of expectations in the last cost term and write

$$\begin{aligned} C = {} & P_0\left\{\int_\mathcal{A} C_{00}pr(\mathbf{g}|0)\mathrm{d}^M g + \int_\mathcal{P} C_{10}[\hat{\theta}(\mathbf{g})]pr(\mathbf{g}|0)\mathrm{d}^M g\right\} \\ & + P_1\left\{\int_\mathcal{A}\langle C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\mathrm{d}^M g\right. \\ & \left. + \int_\mathcal{P}\langle C_{11}[\hat{\theta}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\mathrm{d}^M g\right\}. \end{aligned} \tag{3}$$

The notation $\boldsymbol{\theta}|\mathbf{g}$ indicates that the posterior PDF $pr(\boldsymbol{\theta}|\mathbf{g})$ is being used to compute the expectation. By rearranging terms, we have the risk reduced to two integrals:

$$\begin{aligned} C = {} & \int_\mathcal{A}\{P_0C_{00}pr(\mathbf{g}|0) + P_1\langle C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\}\mathrm{d}^M g \\ & + \int_\mathcal{P}\{P_0C_{10}[\hat{\theta}(\mathbf{g})]pr(\mathbf{g}|0) \\ & + P_1\langle C_{11}[\hat{\theta}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\}\mathrm{d}^M g. \end{aligned} \tag{4}$$

We will now proceed with some mathematical manipulations of this expression that will lead to the ideal classifier for any estimator $\hat{\boldsymbol{\theta}}(\mathbf{g})$. Then, we will find the ideal estimator to complete the ideal observer for the joint task.

We start by defining a quantity B that does not depend on the classifier or the estimator:

$$\int_{\mathcal{D}}\{P_0C_{00}pr(\mathbf{g}|0)+P_1\langle C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\}\mathrm{d}^Mg$$
$$=P_0C_{00}+P_1\langle C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}}=B. \tag{5}$$

We may now write risk as $C=\tilde{C}+B$ with

$$\tilde{C}=-\int_{\mathcal{P}}\{P_0C_{00}pr(\mathbf{g}|0)+P_1\langle C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\}\mathrm{d}^Mg$$
$$+\int_{\mathcal{P}}\{P_0C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g})]pr(\mathbf{g}|0)$$
$$+P_1\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\}\mathrm{d}^Mg. \tag{6}$$

Since B does not depend on the classifier, we want to choose the set $\mathcal{P}$ to minimize $\tilde{C}$. This implies that we must choose the set $\mathcal{P}$ to be exactly the set of data vectors that satisfy the inequality:

$$P_0C_{10}[\hat{\theta}(\mathbf{g})]pr(\mathbf{g}|0)+P_1\langle C_{11}[\hat{\theta}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)$$
$$\leq P_0C_{00}pr(\mathbf{g}|0)+P_1\langle C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1). \tag{7}$$

By using the assumption that $C_{10}(\hat{\theta})\geq C_{00}$, we may convert this inequality to

$$\frac{\langle C_{01}[\boldsymbol{\theta}]-C_{11}[\hat{\theta}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}}{C_{10}[\hat{\theta}(\mathbf{g})]-C_{00}}\Lambda(\mathbf{g})\geq\frac{P_0}{P_1}, \tag{8}$$

where $\Lambda(\mathbf{g})=pr(\mathbf{g}|1)/pr(\mathbf{g}|0)$ is the likelihood ratio. This is the only point where the condition $C_{10}(\hat{\theta})\geq C_{00}$ is used. The likelihood ratio, or any monotonic transformation of it, is the ideal-observer test statistic for the pure detection task. With the obvious definitions for $\beta(\mathbf{g})$ and $\alpha(\mathbf{g})$ we have, for the joint task, the test statistic $t(\mathbf{g})$ and threshold τ such that

$$t(\mathbf{g})=\frac{\beta(\mathbf{g})}{\alpha(\mathbf{g})}\Lambda(\mathbf{g})\geq\frac{P_0}{P_1}=\tau \tag{9}$$

is required for the signal to be declared present. Note that it is very unlikely that $t(\mathbf{g})$ is the result of a monotonic transformation of $\Lambda(\mathbf{g})$. Thus, the ideal observer for the joint task is giving up some performance on the detection task in order to minimize the overall risk.

By combining the two terms in the risk expression in Eq. (2) that depend on the estimator, we find that the ideal estimator is given by

$$\hat{\boldsymbol{\theta}}(\mathbf{g})=\arg\min_{\boldsymbol{\theta}'}\{P_0C_{10}[\boldsymbol{\theta}']pr(\mathbf{g}|0)$$
$$+P_1\langle C_{11}[\boldsymbol{\theta}',\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\}, \tag{10}$$

since this will minimize the combined integral over $\mathcal{P}$ in the risk no matter what $\mathcal{P}$ actually is. In the absence of a need to classify along with estimate, the ideal estimator would have $P_0=0$. Thus, the ideal observer for the joint task is losing some estimation performance compared to the ideal observer for a pure estimation task. The ideal estimator for the joint task can also be written as

$$\hat{\boldsymbol{\theta}}(\mathbf{g},\tau)=\arg\min_{\boldsymbol{\theta}'}\left\{\tau C_{10}\frac{\Delta y}{\Delta x}[\boldsymbol{\theta}']pr(\mathbf{g}|0)\right.$$
$$\left.+\langle C_{11}[\boldsymbol{\theta}',\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}pr(\mathbf{g}|1)\right\}, \tag{11}$$

where we are now indicating that this estimator depends on the quantity τ used in the classification task. Since the test statistic depends on the estimator, we must now write for the signal present cases

$$t(\mathbf{g},\tau)=\frac{\beta(\mathbf{g},\tau)}{\alpha(\mathbf{g},\tau)}\Lambda(\mathbf{g})\geq\frac{P_0}{P_1}=\tau, \tag{12}$$

where

$$\alpha(\mathbf{g},\tau)=C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g},\tau)]-C_{00} \tag{13}$$

and

$$\beta(\mathbf{g},\tau)=\langle C_{01}[\boldsymbol{\theta}]-C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g},\tau),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}. \tag{14}$$

This dependence of the test statistic and estimator on τ complicates the definition for a corresponding EROC curve. We will address this problem in the next section.

The test statistic and estimator derived here are different than those used by the ideal observer for standard EROC analysis. We get that observer by making $C_{10}[\boldsymbol{\theta}']$ and $C_{01}[\boldsymbol{\theta}]$ constants C_{10} and C_{01}, respectively, and by defining a utility function for the estimator as

$$U[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]=\frac{C_{01}-C_{11}[\boldsymbol{\theta}(\mathbf{g}),\boldsymbol{\theta}]}{C_{10}-C_{00}}. \tag{15}$$

Thus, standard EROC analysis is equivalent to a special case of the risk-based EROC analysis described in this paper. Note that these assumptions also remove the dependence of the estimator, and hence the test statistic, on the threshold. This makes it relatively easy to define the standard EROC curve in analogy with the standard ROC curve. In Ref. 3, the EROC curve problem is generalized using utility instead of cost and an ideal observer equivalent to the one derived here is presented. However, at that point, some assumptions are made about the structure of the utility functions that reduces their generality but permits the definition of EROC-type curves that coincide with curves used in the literature, such as FROC and AFROC curves. In particular, one of the assumptions is that the FN cost or utility does not depend on the true vector $\boldsymbol{\theta}$. We are taking an alternative approach that leaves the cost functions completely general and defining an EROC curve based on them. In Ref. 6, the ideal observer for the cost matrix we are using here is derived for the FROC and AFROC problems where constraints are placed on the cost functions. The resulting ideal observer is equivalent to the one derived here, although the expressions look more complicated than the ones above. The cost constraints in Ref. 6 are needed to make sure the risk minimizing observer also maximizes the FROC and AFROC curves. In both Refs. 3 and 6, the relation between the ideal classifier for the joint task and the likelihood ratio is not as obvious as it is in Eqs. (8) and (9) above because they do not make use of the posterior distribution on the parameter vector. This relationship, in turn, makes it clear, as noted above, that the ideal observer for the joint task is giving up detection performance, compared to the ideal observer on the pure detection task, in order to improve estimation performance. For these reasons, we have provided the derivation above for the ideal observer on the joint task.

2.2 Risk-Based EROC Curve for Application I

Now we will define the RB-EROC curve for the ideal observer in application I. Using the fact that $\text{step}(-x) = 1 - \text{step}(x)$, we can write the risk in the form

$$C(\tau) = P_0 \langle C_{00} + \text{step}[t(\mathbf{g},\tau) - \tau]\alpha(\mathbf{g},\tau)\rangle_{\mathbf{g}|0} + P_1 \langle\langle C_{11}[\hat{\theta}(\mathbf{g},\tau),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} + \text{step}[\tau - t(\mathbf{g},\tau)]\beta(\mathbf{g},\tau)\rangle_{\mathbf{g}|1}. \tag{16}$$

The first expectation involves a differential cost for false positives so we write

$$x(\tau) = \langle C_{00} + \text{step}[t(\mathbf{g},\tau) - \tau]\alpha(\mathbf{g},\tau)\rangle_{\mathbf{g}|0}. \tag{17}$$

The second expectation contains a differential cost for false negatives and we write

$$y(\tau) = \langle\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g},\tau),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} + \text{step}[\tau - t(\mathbf{g},\tau)]\beta(\mathbf{g},\tau)\rangle_{\mathbf{g}|1}. \tag{18}$$

The ideal observer RB-EROC curve for application I is a plot of $y(\tau)$ versus $x(\tau)$ as τ is varied from 0 to ∞. As $\tau \to \infty$, we have $x(\tau)$ is decreasing toward C_{00} and $y(\tau)$ is increasing toward $\langle C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}}$. At $\tau = 0$ we have

$$x(0) = \langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g},0)]\rangle_{\mathbf{g}|0} > C_{00},$$
$$y(0) = \langle\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g},0),\boldsymbol{\theta}]\rangle_{\mathbf{g}|\boldsymbol{\theta}}\rangle_{\boldsymbol{\theta}} C_{01}[\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}}. \tag{19}$$

Thus, the curve is decreasing from left to right as τ decreases.

Since the function $x(\tau)$ is monotonic, each point on the ideal-observer RB-EROC curve determines a value for $\tau = P_0/P_1$. Unfortunately, determining τ from the point on the curve is not straightforward. This is due to the complications introduced by the dependence of the estimator on τ. Nevertheless we have

$$C(\tau) = P_0 x(\tau) + P_1 y(\tau), \tag{20}$$

so the ideal-observer RB-EROC curve determines the minimum risk for all values of the prior probabilities.

The ideal observer in the previous section minimizes, over all observers, $y(\tau)$ for any given value of $x(\tau)$ and therefore has the lowest RB-EROC curve. We can show this as follows. Suppose some other observer has a smaller y-coordinate y_0 at a given x-coordinate x_0 on its RB-EROC curve than the ideal observer. Since $x(\tau)$ is a monotonic function of τ, the given x-coordinate determines a value for τ such that $x(\tau) = x_0$. This τ then determines $y(\tau)$ and P_0 and P_1. Then, the risk $P_0 x_0 + P_1 y_0$ would be less than $P_0 x(\tau) + P_1 y(\tau)$, which is a contradiction. Thus a version of the Newmann–Pearson lemma holds for the RB-EROC curve. This argument is easily reversed to show that minimizing the RB-EROC curve also minimizes the risk.

We can get the standard EROC curve by setting $C_{00} = 0$ and $C_{10} = C_{01} = 1$. Then, the RB-EROC curve for this application is an upside down EROC curve with the utility function $U[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}] = 1 - C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]$.

2.3 Area Under the RB-EROC Curve for Application I

The RB-EROC curve can be plotted for any observer on the joint task that uses a test statistic with a threshold for the detection component. For most conventional observers, the estimator will not depend on the prior probabilities for the two classes, and hence α and β will also be independent of these probabilities. This is not an optimum strategy in the sense of minimizing the risk, but it is an understandable one in view of the complications involved in calculating the ideal observer. In this case, τ is simply the threshold being used for the classification task and is not necessarily connected to the prior probabilities on the classes. We then have, for the risk,

$$C(\tau) = P_0\{\langle C_{00} + \text{step}[t(\mathbf{g}) - \tau]\alpha(\mathbf{g})\rangle_{\mathbf{g}|0}\} + P_1\{\langle\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} + \text{step}[\tau - t(\mathbf{g})]\beta(\mathbf{g})\rangle_{\mathbf{g}|1}\}, \tag{21}$$

where

$$\alpha(\mathbf{g}) = C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g})] - C_{00} \tag{22}$$

and

$$\beta(\mathbf{g}) = \langle C_{01}[\boldsymbol{\theta}] - C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}. \tag{23}$$

For the RB-EROC curve, we define

$$x(\tau) = \langle C_{00} + \text{step}[t(\mathbf{g}) - \tau]\alpha(\mathbf{g})\rangle_{\mathbf{g}|0} \tag{24}$$

and

$$y(\tau) = \langle\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} + \text{step}[\tau - t(\mathbf{g})]\beta(\mathbf{g})\rangle_{\mathbf{g}|1}. \tag{25}$$

For the derivative of $x(\tau)$, we then have

$$\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau} = -\langle\delta[t(\mathbf{g}') - \tau]\alpha(\mathbf{g}')\rangle_{\mathbf{g}'|0}. \tag{26}$$

Note the negative sign here since $x(\tau)$ is a decreasing function of τ. We take this negative sign into account when we define the area under the RB-EROC curve as

$$A = -\int_0^\infty y(\tau)\mathrm{d}x(\tau). \tag{27}$$

We will not use the acronym AUC (area under the curve) for this integral since this is used for standard ROC and EROC curves that increase from left to right and are maximized by the corresponding ideal observers. By using the delta function to perform the integration over τ, we arrive at

$$A = \langle\langle \text{step}[t(\mathbf{g}') - t(\mathbf{g})]\alpha(\mathbf{g}')\beta(\mathbf{g})\rangle_{\mathbf{g}'|0}\rangle_{\mathbf{g}|1}. \tag{28}$$

This can be interpreted as the result of a 2AFC test where the observer pays the penalty $\alpha(\mathbf{g}')\beta(\mathbf{g})$ when it misclassifies $\mathbf{g}'$ as coming from the signal present class and $\mathbf{g}$ from the signal absent class. Unfortunately, there is no simple expression like this for the area under the ideal-observer RB-EROC curve. This again is due the dependence on τ of the estimator in this case.

2.4 *Slope of the RB-EROC Curve for Application I*

If we examine the derivation of the ideal observer test statistic given above, we can see that, in a situation where we are using an estimator that does not depend on τ, there is a test statistic and threshold that minimizes the risk. The corresponding observer is suboptimal since it is not using the ideal estimator for the joint task but is nevertheless of interest since, as we have seen in the previous section, the area under the RB-EROC curve for such an observer is easy to estimate from a 2AFC test. Therefore, in this section, we again use $\tau = P_0/P_1$ and we have

$$\frac{\mathrm{d}y(\tau)}{\mathrm{d}x(\tau)} = \frac{\langle \delta[\tau - t(\mathbf{g})]\beta(\mathbf{g})\rangle_{\mathbf{g}|1}}{-\langle \delta[t(\mathbf{g}') - \tau]\alpha(\mathbf{g}')\rangle_{\mathbf{g}'|0}}. \tag{29}$$

Now we use the fact that, on the hypersurface defined by $\tau - t(\mathbf{g}) = 0$, we have $\beta(\mathbf{g})pr(\mathbf{g}|1) = \tau\alpha(\mathbf{g})pr(\mathbf{g}|0)$. This fact then gives us

$$\frac{\mathrm{d}y(\tau)}{\mathrm{d}x(\tau)} = -\tau. \tag{30}$$

For these observers on the joint task, the RB-EROC curve then provides all of the information we need to compute the corresponding risk for all values of the prior probabilities.

3 Example for Application I

As an example for application I, we consider quadratic cost functions:

$$\begin{bmatrix} C_{00} & C_{01}[\boldsymbol{\theta}] \\ C_{10}[\hat{\theta}] & C_{11}[\hat{\theta},\boldsymbol{\theta}] \end{bmatrix} = \begin{bmatrix} 0 & c_{01}\|\boldsymbol{\theta}\|^2 \\ c_{10}\|\hat{\theta}\|^2 & c_{11}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \end{bmatrix}. \tag{31}$$

We choose these cost functions because they are the simplest ones to use analytically and will give us some idea of how this whole scheme works out in practice. Realistic cost functions would undoubtedly be more complicated in order to take into account the consequences of each observer outcome.

The estimator used by the ideal observer on the joint task is now given by

$$\hat{\boldsymbol{\theta}}(\mathbf{g},\tau) = \arg\min_{\boldsymbol{\theta}'}\{\tau c_{10}\|\boldsymbol{\theta}'\|^2 pr(\mathbf{g}|0) + c_{11}\langle\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|^2\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}. \tag{32}$$

We define $\bar{\boldsymbol{\theta}}_p(\mathbf{g})$ to be the mean of the posterior PDF $pr(\boldsymbol{\theta}|\mathbf{g})$, and $\mathbf{K}_p(\mathbf{g})$ to be the covariance matrix for this distribution. We will only need the trace of this covariance matrix so we define the total variance of the posterior PDF as $\sigma_p^2(\mathbf{g}) = \mathrm{tr}[\mathbf{K}_p(\mathbf{g})]$. Now we have

$$\langle\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|^2\rangle_{\boldsymbol{\theta}|\mathbf{g}} = \|\boldsymbol{\theta}'\|^2 - 2\bar{\boldsymbol{\theta}}_p(\mathbf{g})\cdot\boldsymbol{\theta}' + \sigma_p^2(\mathbf{g}) + \|\bar{\boldsymbol{\theta}}_p(\mathbf{g})\|^2. \tag{33}$$

Suppressing some dependencies on $\mathbf{g}$ and $\tau$for the moment, we can now write the estimate as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}'}\{A\|\boldsymbol{\theta}'\|^2 - 2B\bar{\theta}_p\cdot\boldsymbol{\theta}' + C\}, \tag{34}$$

with $A = \tau c_{10}pr(\mathbf{g}|0) + c_{11}P_1 pr(\mathbf{g}|1)$, $B = c_{11}P_1 pr(\mathbf{g}|1)$, and $C = c_{11}(\sigma_p^2 + \|\bar{\boldsymbol{\theta}}_p\|^2)P_1 pr(\mathbf{g}|)$. Taking the gradient, we have $A\hat{\boldsymbol{\theta}} - B\bar{\boldsymbol{\theta}}_p = \mathbf{0}$. Therefore, the ideal estimator we are looking for is given by

$$\hat{\boldsymbol{\theta}}(\mathbf{g},\tau) = \left[\frac{c_{11}pr(\mathbf{g}|1)}{\tau c_{10}pr(\mathbf{g}|0) + c_{11}pr(\mathbf{g}|1)}\right]\bar{\theta}_p(\mathbf{g}). \tag{35}$$

In terms of the likelihood ratio, we may write this estimator as

$$\hat{\boldsymbol{\theta}}(\mathbf{g},\tau) = \left[\frac{c_{11}\Lambda(\mathbf{g})}{\tau c_{10} + c_{11}\Lambda(\mathbf{g})}\right]\bar{\boldsymbol{\theta}}_p(\mathbf{g}) = \gamma(\mathbf{g},\tau)\bar{\boldsymbol{\theta}}_p(\mathbf{g}), \tag{36}$$

where we are using this equation to define the function $\gamma(\mathbf{g},\tau)$. This shows that, for relatively large values of the likelihood ratio, when the observer is more likely to declare that the signal is present, the estimator is approximately the same as the posterior mean estimator, which is the ideal estimator for the pure estimation task with the quadratic penalty function that we have here. On the other hand, when the likelihood ratio is relatively small, the observer is more likely to declare that the signal is absent. In this case, the ideal estimator for the joint tasks multiplies the posterior mean estimator by a small scalar, which reduces the false positive cost function.

The ideal classifier for this example classifies the data as signal present if

$$\frac{\langle c_{01}\|\boldsymbol{\theta}\|^2 - c_{11}\|\hat{\theta}(\mathbf{g},\tau) - \boldsymbol{\theta}\|^2\rangle_{\boldsymbol{\theta}|\mathbf{g}}}{c_{10}\|\hat{\boldsymbol{\theta}}(\mathbf{g},\tau)\|^2}\Lambda(\mathbf{g}) \geq \frac{P_0}{P_1} = \tau. \tag{37}$$

Performing the expectation in the numerator results in the classifier inequality

$$\frac{\beta(\mathbf{g},\tau)}{\alpha(\mathbf{g},\tau)}\Lambda(\mathbf{g}) \geq \tau, \tag{38}$$

with

$$\beta(\mathbf{g},\tau) = (c_{01} - c_{11})[\sigma_p^2(\mathbf{g}) + \|\bar{\boldsymbol{\theta}}_p(\mathbf{g})\|^2] + c_{11}\gamma(\mathbf{g},\tau)\|\bar{\boldsymbol{\theta}}_p(\mathbf{g})\|^2[2 - \gamma(\mathbf{g},\tau)] \tag{39}$$

and

$$\alpha(\mathbf{g}) = c_{10}\gamma^2(\mathbf{g},\tau)\|\bar{\boldsymbol{\theta}}_p(\mathbf{g})\|^2. \tag{40}$$

Thus, the posterior mean and total variance are all that is needed from the posterior PDF to formulate the ideal observer for this example.

4 Application II: Unconditional Estimation

In this application, the observer must estimate a parameter vector for the signal present and the signal absent classifications. Therefore, the estimation task is not conditional on the outcome of the detection task. The dimension of the parameter vector $\boldsymbol{\theta}$ for the signal present class may be different than the parameter vector $\boldsymbol{\phi}$ for the signal absent class. For example, the two classes may represent two different signals that are parameterized by vectors of different dimensions. In medical imaging, the two classes may represent two different medical conditions that could give rise to the presented symptoms. We may then want to estimate different parameter vectors for the two conditions in order to proceed with treatment. The PDFs for the two classes are given by

$$pr(\mathbf{g}|0) = \int_{\Theta} pr(\mathbf{g}|\boldsymbol{\phi})pr(\boldsymbol{\phi})\mathrm{d}^K\phi \tag{41}$$

and

$$pr(\mathbf{g}|1) = \int_\Theta pr(\mathbf{g}|\boldsymbol{\theta})pr(\boldsymbol{\theta})\mathrm{d}^L\theta. \tag{42}$$

The cost functions in the cost matrix will now depend on two vector variables, the estimate produced for the given classification, and the true parameter vector for the object being classified. The cost matrix is therefore written as

$$\begin{bmatrix} C_{00}[\hat{\boldsymbol{\phi}},\boldsymbol{\phi}] & C_{01}[\hat{\boldsymbol{\phi}},\boldsymbol{\phi}] \\ C_{10}[\hat{\boldsymbol{\theta}},\boldsymbol{\phi}] & C_{11}[\hat{\boldsymbol{\theta}},\boldsymbol{\theta}] \end{bmatrix}.$$

We will make the simplifying assumptions

$$C_{00}[\hat{\boldsymbol{\phi}},\boldsymbol{\phi}]_{10}[\hat{\boldsymbol{\theta}},\boldsymbol{\theta}] \tag{43}$$

and

$$C_{11}[\hat{\boldsymbol{\theta}},\boldsymbol{\theta}]_{01}[\hat{\boldsymbol{\phi}},\boldsymbol{\theta}] \tag{44}$$

In other words, the cost of a misclassification is greater than the cost for the corresponding correct classification. The average cost or risk C for the combined task is then given by

$$\begin{aligned} C = P_0 \Bigg\{ & \left\langle \int_{\mathcal{A}} C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]pr(\mathbf{g}|\boldsymbol{\phi})d^M g \right\rangle_{\boldsymbol{\phi}} \\ & + \left\langle \int_{\mathcal{P}} C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\phi}]pr(\mathbf{g}|\boldsymbol{\phi})d^M g \right\rangle_{\boldsymbol{\phi}} \Bigg\} \\ & + P_1 \Bigg\{ \left\langle \int_{\mathcal{A}} C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\theta}]pr(\mathbf{g}|\boldsymbol{\theta})d^M g \right\rangle_{\boldsymbol{\phi}} \\ & + \left\langle \int_{\mathcal{P}} C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]pr(\mathbf{g}|\boldsymbol{\phi})d^M g \right\rangle_{\boldsymbol{\phi}} \Bigg\}. \end{aligned} \tag{45}$$

The notation here is the same as in application I in terms of the test statistic, threshold, and estimators. The calculations that follow are very similar to those in application I, they just look more complex due to the added complexity of the cost matrix. The conclusions that we arrive at are straightforward generalizations of those in application I.

4.1 *Ideal Observer for Application II*

For the ideal observer, we make use of posterior PDFs for both parameter vectors to write the risk as

$$\begin{aligned} C = & \int_{\mathcal{A}} \{P_0\langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) \\ & + P_1\langle C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}\mathrm{d}^M g \\ & + \int_{\mathcal{P}} \{P_0\langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) \\ & + P_1\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}\mathrm{d}^M g. \end{aligned} \tag{46}$$

We now define a constant B, which does not depend on the classifier, as

$$\begin{aligned} & \int_{\mathcal{D}} \{P_0\langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) \\ & + P_1\langle C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}\mathrm{d}^M g \\ & = P_0\langle\langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}}\rangle_{\mathbf{g}|0} + P_1\langle\langle C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}\rangle_{\mathbf{g}|1} = B. \end{aligned} \tag{47}$$

This allows us to write $C = \tilde{C} + B$ and choose the classifier to minimize

$$\begin{aligned} \tilde{C} = & -\int_{\mathcal{P}} \{P_0\langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) \\ & + P_1\langle C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}\mathrm{d}^M g \\ & + \int_{\mathcal{P}} \{P_0\langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) \\ & + P_1\langle C_{11}[\boldsymbol{\theta}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}\mathrm{d}^M g. \end{aligned} \tag{48}$$

The decision rule is therefore to classify $\mathbf{g}$ as signal present if and only if

$$\begin{aligned} & P_0\langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) + P_1\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1) \\ & \le P_0\langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) + P_1\langle C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1), \end{aligned} \tag{49}$$

and otherwise classify $\mathbf{g}$ as signal absent. Using our assumptions about the cost functions, we can write the decision rule for the signal present decision as

$$\frac{\langle C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} - \langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}}{\langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} - \langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}}}\Lambda(\mathbf{g}) \ge \frac{P_0}{P_1}. \tag{50}$$

With the obvious definitions for $\mu(\mathbf{g})$ and $\lambda(\mathbf{g})$, we will write this inequality as

$$\frac{\mu(\mathbf{g})}{\lambda(\mathbf{g})}\Lambda(\mathbf{g}) \ge \frac{P_0}{P_1}. \tag{51}$$

This decision rule is used by the ideal observer for the joint task in this application. Note that the notation $\lambda(\mathbf{g})$ is often used for the natural logarithm of the likelihood ratio, but we are not using the log-likelihood in this work so no confusion should arise.

Isolating those parts of the risk that depend on the respective estimators, we find that the ideal estimators for the joint task are given by

$$\begin{aligned} \hat{\boldsymbol{\phi}}(\mathbf{g},\tau) = \arg\min_{\boldsymbol{\phi}'} \{ & \tau\langle C_{00}[\boldsymbol{\phi}',\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) \\ & + \langle C_{01}[\boldsymbol{\phi}',\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\} \end{aligned} \tag{52}$$

and

$$\begin{aligned} \hat{\boldsymbol{\theta}}(\mathbf{g},\tau) = \arg\min_{\boldsymbol{\theta}'} \{ & \tau\langle C_{10}[\boldsymbol{\theta}',\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) \\ & + \langle C_{11}[\boldsymbol{\theta}',\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}. \end{aligned} \tag{53}$$

Once again the estimators depend on the threshold τ. This forces us to write the classification rule as

$$\frac{\mu(\mathbf{g},\tau)}{\lambda(\mathbf{g},\tau)}\Lambda(\mathbf{g}) \geq \frac{P_0}{P_1} = \tau \tag{54}$$

with

$$\mu(\mathbf{g},\tau) = \langle C_{01}[\hat{\boldsymbol{\phi}}(\mathbf{g},\tau),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} - \langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g},\tau),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} \tag{55}$$

and

$$\lambda(\mathbf{g},\tau) = \langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g},\tau),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} - \langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g},\tau),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}}. \tag{56}$$

As in application I, some performance on the pure classification and estimation tasks is given up by the ideal observer for the joint task.

4.2 *Risk-Based EROC Curve for Application II*

As in application I, the risk can be rewritten as follows:

$$C = P_0\langle\langle \mathrm{C}_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} + \mathrm{step}[t(\mathbf{g}') - \tau]\lambda(\mathbf{g},\tau)\rangle_{\mathbf{g}|0}$$
$$+ P_1\langle\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} + \mathrm{step}[\tau - t(\mathbf{g})]\mu(\mathbf{g},\tau)\rangle_{\mathbf{g}|1}. \tag{57}$$

This leads to a function related to the differential cost of the false positive outcomes

$$x(\tau) = \langle\langle \mathrm{C}_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g}),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} + \mathrm{step}[t(\mathbf{g}') - \tau]\lambda(\mathbf{g},\tau)\rangle_{\mathbf{g}|0}. \tag{58}$$

We also have a function related to the differential cost of the false negative outcomes

$$y(\tau) = \langle\langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g}),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}} + \mathrm{step}[\tau - t(\mathbf{g})]\mu(\mathbf{g},\tau)\rangle_{\mathbf{g}|1}. \tag{59}$$

The RB-EROC curve for this application is a plot of $y(\tau)$ versus $x(\tau)$ as the threshold τ is varied from 0 to ∞. Just as in application I, this curve will decrease from left to right. At the left most point where $\tau \to \infty$, we have a maximum with

$$x(\tau) \to \langle C_{00}[\hat{\boldsymbol{\phi}}(\mathbf{g},\infty),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} \tag{60}$$

and

$$y(\tau) \to \langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g},\infty),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}}. \tag{61}$$

At the right most point $\tau = 0$, with

$$x(0) = \langle C_{10}[\hat{\boldsymbol{\theta}}(\mathbf{g},0),\boldsymbol{\phi}]\rangle_{\boldsymbol{\phi}|\mathbf{g}} \tag{62}$$

and

$$y(0) = \langle C_{11}[\hat{\boldsymbol{\theta}}(\mathbf{g},0),\boldsymbol{\theta}]\rangle_{\boldsymbol{\theta}|\mathbf{g}}. \tag{63}$$

This curve can be plotted for any observer in this application. The ideal observer in the previous section again minimizes, over all observers, $y(\tau)$ for any given value of $x(\tau)$, and therefore has the lowest RB-EROC curve. This Neymann–Pearson type result can be proved the same way as for application I, since we still have

$$C(\tau) = P_0 x(\tau) + P_1 y(\tau), \tag{64}$$

for all τ.

4.3 *Area Under the RB-EROC Curve for Application II*

There is no simple formula for the area under the ideal-observer RB-EROC curve for application II. However, if we are using suboptimal estimators that are independent of the prior class probabilities then we may proceed as we did in application I. We start with

$$\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau} = \langle\delta[t(\mathbf{g}') - \tau]\lambda(\mathbf{g}')\rangle_{\mathbf{g}'|0}. \tag{65}$$

Then, we define the area as

$$A = -\int_0^\infty y(\tau)\mathrm{d}x(\tau). \tag{66}$$

This results in the double expectation

$$A = \langle\langle \mathrm{step}[t(\mathbf{g}') - t(\mathbf{g})]\lambda(\mathbf{g}')\mu(\mathbf{g})\rangle_{\mathbf{g}'|0}\rangle_{\mathbf{g}|1}. \tag{67}$$

As with application I, this formula may be interpreted as the outcome of a 2AFC test with the penalty $\lambda(\mathbf{g}')\mu(\mathbf{g})$ when the images are misclassified.

4.4 *Slope of the RB-EROC Curve for Application II*

As in application I, if the estimator is independent of the prior class probabilities, then the derivation above provides the optimal decision strategy for this estimator. The slope calculation for the RB-EROC curve in this calculation follows steps similar to those in application I. The derivatives of the horizontal and vertical coordinates are

$$\frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau} = -\langle\delta[t(\mathbf{g}') - \tau]\lambda(\mathbf{g}')\rangle_{\mathbf{g}'|0} \tag{68}$$

and

$$\frac{\mathrm{d}y(\tau)}{\mathrm{d}t_0} = \langle\delta[\tau - t(\mathbf{g})]\mu(\mathbf{g})\rangle_{\mathbf{g}|1}. \tag{69}$$

The delta function confines the integration to the hypersurface defined by $\tau = t(\mathbf{g})$. When we use the definition of $t(\mathbf{g})$, we have

$$\frac{\mathrm{d}y(\tau)}{\mathrm{d}\tau} = \tau\langle\delta[\tau - t(\mathbf{g})]\lambda(\mathbf{g})\rangle_{\mathbf{g}|1}. \tag{70}$$

The end result is the same as in application I

$$\frac{\mathrm{d}y(\tau)}{\mathrm{d}x(\tau)} = -\tau. \tag{71}$$

For these observers on the joint task, the RB-EROC curve then provides all of the information we need to compute the corresponding risk for all values of the prior probabilities.

5 Example for Application II

For an example, we will extend the quadratic cost functions in the application I example to application II. Therefore, we have a cost matrix

$$\begin{bmatrix} C_{00}[\hat{\phi}, \boldsymbol{\phi}] & C_{01}[\hat{\phi}, \boldsymbol{\theta}] \\ C_{10}[\hat{\theta}, \boldsymbol{\phi}] & C_{11}[\hat{\theta}, \boldsymbol{\theta}] \end{bmatrix} = \begin{bmatrix} c_{00}\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\|^2 & c_{01}(\|\hat{\boldsymbol{\phi}}\|^2 + \|\boldsymbol{\theta}\|^2) \\ c_{10}(\|\hat{\boldsymbol{\theta}}\|^2 + \|\boldsymbol{\phi}\|^2) & c_{11}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \end{bmatrix}. \tag{72}$$

The estimators are then defined by

$$\hat{\boldsymbol{\phi}}(\mathbf{g}, \tau) = \arg\min_{\boldsymbol{\phi}'}\{c_{00}\tau\langle\|\boldsymbol{\phi}' - \boldsymbol{\phi}\|^2\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) + c_{01}\langle(\|\boldsymbol{\phi}'\|^2 + \|\boldsymbol{\theta}\|^2)\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\} \tag{73}$$

and

$$\hat{\theta}(\mathbf{g}, \tau) = \arg\min_{\boldsymbol{\theta}'}\{c_{10}\tau\langle\|\boldsymbol{\theta}'\|^2 + \|\boldsymbol{\phi}\|^2\rangle_{\boldsymbol{\phi}|\mathbf{g}} pr(\mathbf{g}|0) + c_{11}\langle\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|^2\rangle_{\boldsymbol{\theta}|\mathbf{g}} pr(\mathbf{g}|1)\}. \tag{74}$$

As in application I, the resulting estimators are scalar multiples of the corresponding posterior mean estimators, which are ideal estimators if there was no classification involved. The ideal application II estimators are

$$\hat{\boldsymbol{\theta}}(\mathbf{g}, \tau) = \left[\frac{c_{11}\Lambda(\mathbf{g})}{c_{10}\tau + c_{11}\Lambda(\mathbf{g})}\right]\bar{\boldsymbol{\theta}}_p(\mathbf{g}) = \gamma(\mathbf{g}, \tau)\bar{\boldsymbol{\theta}}_p(\mathbf{g}) \tag{75}$$

and

$$\hat{\boldsymbol{\phi}}(\mathbf{g}, \tau) = \left[\frac{c_{00}\tau}{c_{01}\Lambda(\mathbf{g}) + c_{00}\tau}\right]\bar{\phi}_p(\mathbf{g}) = \nu(\mathbf{g}, \tau)\bar{\boldsymbol{\phi}}_p(\mathbf{g}). \tag{76}$$

The classifier declares signal present when

$$\frac{\langle c_{01}(\|\hat{\boldsymbol{\phi}}(\mathbf{g},\tau)\|^2 + \|\boldsymbol{\theta}\|^2)\rangle_{\boldsymbol{\theta}|\mathbf{g}} - \langle c_{11}\|\boldsymbol{\theta}(\mathbf{g},\tau) - \boldsymbol{\theta}\|^2\rangle_{\boldsymbol{\theta}|\mathbf{g}}}{\langle c_{10}(\|\hat{\boldsymbol{\theta}}(\mathbf{g},\tau)\|^2 + \|\boldsymbol{\phi}\|^2)\rangle_{\boldsymbol{\phi}|\mathbf{g}} - \langle c_{00}\|\hat{\boldsymbol{\phi}}(\mathbf{g},\tau) - \boldsymbol{\phi}\|^2\rangle_{\boldsymbol{\phi}|\mathbf{g}}}\Lambda(\mathbf{g}) \geq \tau. \tag{77}$$

We can write this inequality as

$$\frac{\mu(\mathbf{g}, \tau)}{\lambda(\mathbf{g}, \tau)}\Lambda(\mathbf{g}) \geq \tau, \tag{78}$$

with

$$\mu(\mathbf{g}, \tau) = (c_{01} - c_{11})[\sigma_p^2(\mathbf{g}) + \|\bar{\boldsymbol{\theta}}_p(\mathbf{g})\|^2] + c_{11}\gamma(\mathbf{g}, \tau)\|\bar{\boldsymbol{\theta}}_p(\mathbf{g})\|^2[2 - \gamma(\mathbf{g}, \tau)] + c_{01}\nu^2(\mathbf{g}, \tau)\|\bar{\boldsymbol{\phi}}_p(\mathbf{g})\|^2 \tag{79}$$

and

$$\lambda(\mathbf{g}, \tau) = (c_{10} - c_{00})[\delta_p^2(\mathbf{g}) + \|\hat{\boldsymbol{\phi}}_p(\mathbf{g})\|^2] + c_{00}\nu(\mathbf{g}, \tau)\|\bar{\boldsymbol{\phi}}_p(\mathbf{g})\|^2[2 - \nu(\mathbf{g})] + c_{10}\gamma^2(\mathbf{g}, \tau)\|\bar{\boldsymbol{\theta}}_p(\mathbf{g})\|^2. \tag{80}$$

As in application I, for this cost matrix, we only need the posterior means and total variances from the posterior PDFs in order to formulate the ideal observer.

6 Conclusion

We have considered a Bayesian risk approach to two types of tasks that combine detection with estimation. In application I, the estimation of a parameter vector only occurs for the signal present classification. In application II, estimation of possibly different parameter vectors occurs with both signal present and signal absent classifications. The cost functions are the most general possible in the sense that all cost functions depend on the relevant estimated and/or true parameter vectors for each classification outcome. In both applications, we found analytical expressions for the estimators and classifiers for the ideal Bayesian observers.

We went on to define the RB-EROC curve for each application, which is minimized by the ideal observer compared to other observers on the same task. One property of the ideal observer for both applications is that the estimator depends on the prior probabilities of the two classes. This complicates the task of developing analytic expressions for the area under RB-EROC curves and the slope of these curves at each point. However, for suboptimal observers using estimators that do not depend on the prior probabilities we developed a 2AFC formula for the area under their RB-EROC curves. For these estimators, we also showed that the slope for the RB-EROC curve when the optimal classifier is used is the negative of the threshold used by the classifier.

The examples for each application used quadratic cost functions since they are the most analytically tractable. The resulting ideal observers only need the posterior mean and the total variance of the posterior PDFs for the parameter vectors in order to perform the classification and estimation tasks. In the Gaussian case, the posterior mean and total variance have analytical formulas, and therefore computation of the RB-EROC curve and its area only require Monte Carlo sampling from the signal present and signal absent distributions. For more general statistics, MCMC methods would be needed.

The extension of these results to three or more classes should be possible, but the complexity of the observers will rapidly increase with the number of classes. This is the subject of ongoing research.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the author.

Acknowledgments

This material was based upon work supported by the National Institutes of Health under R01-EB000803 and P41-EB002035 and by the Department of Homeland Security under Contract No. HSHQDC-14-C-BOOIO.

References

1. E. Clarkson, "The estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks," *J. Opt. Soc. Am. A* **24**, B91–B98 (2007).
2. E. Clarkson, "Estimation ROC curves and their corresponding ideal observers," *Proc. SPIE* **6515**, 651504 (2007).
3. A. Wunderlich, B. Gossens, and C. K. Abbey, "Optimal joint detection and estimation that maximizes ROC-type curves," *IEEE Trans. Med. Imaging* **35**, 2164–2173 (2016).

4. A. Wunderlich and B. Gossens, "Nonparametric estimation receiver operating characteristic analysis for performance evaluation on combined detection and estimation tasks," *J. Med. Imaging* **1**(3), 031002 (2014).
5. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, Wiley-Interscience, Hoboken, New Jersey (2004).
6. P. Khurd, B. Liu, and G. Gindi, "Ideal AFROC and FROC observers," *IEEE Trans. Med. Imaging* **29**, 375–386 (2010).

Eric Clarkson is a professor at the University of Arizona in Tucson, Arizona. He has joint appointments in medical imaging, optical sciences, and applied mathematics. He received his PhD in mathematics and his MS degree in physics from Arizona State University in Tempe, Arizona. He also has an MS degree in optical sciences from the University of Arizona. His undergraduate training was at Rice University in Houston, where he received his BA degree in mathematics, physics, and philosophy.